Optimizing Scan Times of BLE Scanning Systems

(Invited Paper)

John N Daigle*, George A. Humphrey II†, Henry C. Lena V† and Avijit Sarker†
Department of Electrical Engineering,
University of Mississippi
Oxford, MS, USA
Email: *wcdaigle@olemiss.edu, †gahumphr, hclena, asarker@go.olemiss.edu

***Abstract*—In this paper, we present a detailed analysis of Bluetooth Low Energy-based scanning systems wherein the objective is to successfully scan all items of a group of a prescribed size within a prescribed scanning period at a prescribed probability of success, for example, less than one failure in one million. We show that a design based upon independence among collision events fails to achieve the target reliability objectives by roughly an order of magnitude. Our analysis, which is verified through extensive simulation, shows that correlation among the collision events has a major impact upon the scanning time required to successfully scan all the members of a group.**

I. Introduction

This paper discusses optimization of a scanning system based on Bluetooth Low Energy (BLE) [1]. In the particular, items to be scanned are organized into fixed-size groups, and the scanning of each group is called a scanning job (SJ). An SJ fails if the system fails to successfully scan any item of the group. Each member of the group is equipped with a BLE device and broadcasts advertisements at a rate determined by its interadvertisement time, which is the sum of a relatively large fixed period and a small random period as described in [1] and reviewed here. Each fixed-size group is moved into a scanning area to be scanned. In this paper, an approach to specifying both the interadvertisement time that minimizes scanning time, and the required minimum scanning time, while maintaining an SJ failure (SJF) rate less than a prescribed target failure probability is presented. For example, for a group size of 100, which interadvertisement time should be chosen and how long should the group be scanned so that the system fails to successfully scan every item in the group in less than one group out of each 100 thousand groups on average?

An extensive search of the literature has yielded very few papers that have addressed the specific question addressed here. So far as we are able to determine, all of the references in the literature assume that overlap in transmissions result in collisions that result in failure of both advertisements. In addition, all references assume Poisson arrivals, or equivalently, that advertisements arrive at random points in time and that events of successive collisions are independent. The focus of [2] is to find the optimal value for an advertising interval to minimize the time in which all surrounding BLE advertisers are discovered by a scanner, which addresses part of the issue addressed here. However, [2] does not consider the probability of failing to detect all advertisers that are present, which is the main focus of this paper. Methods for improving discovery time [3], [4], [5], and decreasing the likelihood of packet collision have appeared extensively in the literature [6], [7]. Several other of these papers, including [8], concentrate on how packet collision affects energy consumption of BLE systems, especially in applications dealing with smart health-monitoring devices and devices that provide location services [9], [10]. Al Kalaa [11] focuses on maximizing aggregate throughput on the data channels. To the best of our knowledge, there has not been anything in the literature that directly addresses optimizing interadvertisement time to minimize the probability of collision for a given number of advertisers in a given scan period.

The research reported here was supported in part by NSF I/UCRC-1822104.

In section II, we provide a brief overview of the scanning system and develop simple expressions for the optimal inter-advertisement time and minimum scanning period required to achieve a target $P\{\text{SJF}\}$ based on independence among collision events. To the best of our knowledge, these expressions have not previously appeared in the literature. In section III we analyze the results of a detailed simulation of the system and find that the results based on the optimal design under the assumption of independent collision events miss the target by roughly an order of magnitude. In section IV, we consider the impact of dependence among advertisement events upon the collision process and find the impact to be significant. To the best of our knowledge, the results presented in section IV are also new. The analytical results presented in section IV are also examined via simulation. In section V, we reconsider the design problem using the results of our detailed collision analysis and show that the new design is effective in achieving the system design goals. In section VI, we summarize our results and draw conclusions.

II. Preliminary Analysis

Fig. 1 shows the overall organization of the scanning process, which is now explained. For BLE, the 2400 MHz Industrial, Scientific and Medical (ISM) spectrum is organized as a set of 40 2 MHz channels spaced at 2 MHz intervals with channels centered at 2401 to 2479 MHz. The channels at 2402, 2426, and 2480 MHz are indexed 37, 38, and 39, respectively, and are the advertising channels. Scanning is organized into a sequence of scanning intervals, during each of which the scanner scans for advertisements during its scanning window

978-1-7281-6607-0/20/$31.00 ©2020 IEEE

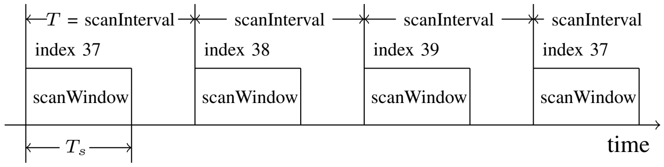

Fig. 1. Illustration of scanning windows and scanning intervals

in one of the advertising channels in sequence. In the general use case, the scanning window can be less than the scanning interval, but in the case of a pure scanning system the scanner scans during the entire scanning window.

Asynchronously from the scanner and independent of all other advertisers, each advertiser is broadcasting advertisements according to its own schedule as shown in Fig. 2. During each advertising cycle, the advertiser broadcasts three advertisements, one in each advertising channel, at fixed time intervals of 10 ms or less. Each advertising event is followed by a relatively long fixed period of radio silence, which is then followed by a random period of silence uniformly distributed over (0, 10) ms. Also, T_{AI} is a parameter that can be adjusted by the user whereas the dead time is typically set to less than 10 ms by the manufacturer, but cannot be set by the user.

The sequence of interadvertisement events is a sequence of independent, identically distributed (IID) random variables over $(T_{\text{AI}}, T_{\text{AI}} + T_D)$, where $T_D = 10$ ms. Thus, the sequence of times at which advertising events occur are a renewal process. Given that the start time of the initial advertising event occurs at a random point in time, the probability that the advertiser is transmitting at one of the frequencies at an arbitrary point in time is given by

$$\rho_{\text{A}} = T_{\text{A}}/\text{E}\left[\tilde{t}_{\text{IA}}\right],$$

where T_{A} is the time required to transmit one advertisement and $\tilde{t}_{\text{IA}} = T_{\text{AI}} + \tilde{t}_{\text{AD}}$ as illustrated in Fig. 2. For the purposes of this analysis, $T_{\text{A}} = 376$ μs, which is the time required to transmit 376 bits at 1 Mb/s as prescribed in [1]. Since the advertisement times of the advertisers are independent, the probability that a second advertiser advertises over the period of advertisement of a tagged advertiser is equal to the probability that the second advertiser's advertisement begins within $\pm T_{\text{A}}$ of the start time of the tagged advertisement. As is commonly assumed in the literature, it is assumed here that overlap in transmission times results in a failed advertisement on the part of both advertisers. Thus, the probability of failure due to a second advertisement is $2\rho_{\text{A}}$.

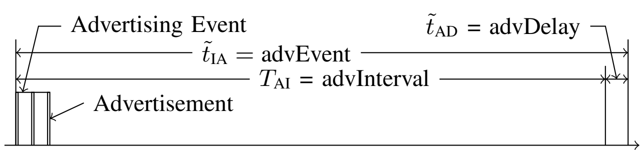

Fig. 2. Advertising events, intervals, and delays (more to scale)

Define $\mathcal{S}_{\text{AA}}$ to be the event of a successful scan for an arbitrary advertisement. Then, since there are N_{A} advertisers and they all transmit independently, the probability that an arbitrary advertisement is successful is given by

$$P\left\{\mathcal{S}_{\text{AA}}\right\} = \left(1 - 2\rho_{\text{A}}\right)^{N_A - 1} \approx e^{-2\rho_{\text{A}}(N_A - 1)}, \quad (1)$$

where the approximation is based on the definition of the mathematical constant e and is very good for large N_A.

Recall that a scanning job (SJ) is the activity of scanning a population of advertisers of a given size over a given period of time, T_S. Each advertiser has its own identity, and the purpose of the SJ is to determine which specific advertisers are in each particular job. If all of the advertisers are scanned successfully at least once during the scanning period, the scanning activity is considered a success; otherwise the scanning job fails. Assuming the events of successful scans are independent and that the advertiser advertises N_S times during an advertising period, we find

$$P\left\{\text{SJF}\right\} = \left(1 - e^{-2\rho_{\text{A}}(N_A - 1)}\right)^{N_S}. \quad (2)$$

In practice, the number of times an advertisement is scanned depends upon the interadevertisement time, $\tilde{t}_{\text{IA}}$, which depends upon the value to which T_{IA} is set and the amount of time selected for scanning. Since the random part of $\tilde{t}_{\text{IA}}$ is very small compared to T_{IA}, we find that $N_S \approx T_S/\text{E}\left[\tilde{t}_{\text{IA}}\right]$. For the purposes of the present analysis, define $N_S = T_S/\text{E}\left[\tilde{t}_{\text{IA}}\right]$ and so that, in general, N_S turns out to be a mixed number. The implication is that during the scanning periods, a fraction of the advertisers would have an additional scan above $\left\lfloor T_S/\text{E}\left[\tilde{t}_{\text{IA}}\right]\right\rfloor$. The resulting scanning job failure probability is

$$P\left\{\text{SJF}\right\} = \left(1 - e^{-2\frac{T_{\text{A}}(N_A - 1)}{\text{E}\left[\tilde{t}_{\text{IA}}\right]}}\right)^{\frac{T_S}{\text{E}\left[\tilde{t}_{\text{IA}}\right]}}. \quad (3)$$

The previous equation has the form

$$P\left\{\text{SJF}\right\} = \left(1 - e^{-ax}\right)^{xT_S}, \quad (4)$$

where $a = 2T_{\text{A}}\left(N_A - 1\right)$ and $x = 1/\text{E}\left[\tilde{t}_{\text{IA}}\right]$, from which it is clear that the value of x, or equivalently $\text{E}\left[\tilde{t}_{\text{IA}}\right]$, that minimizes

$$f(x) = \left(1 - e^{-ax}\right)^x$$

also minimizes $P\left\{\text{SJF}\right\}$ for any scanning period T_S. Additionally, we note that the value of x that minimizes $f(x)$ is the same value that maximizes $\ln f(x)$, the latter resulting from the fact that $f(x) \in (0, 1)$. Thus, we can maximize $\ln f(x) = x \ln\left(1 - e^{-ax}\right)$. To simplify the math, define

$$y = 1 - e^{-ax} \quad \Rightarrow x = -\frac{\ln\left(1 - y\right)}{a}, \quad (5)$$

We then define

$$g(y) = -\frac{1}{a}\ln\left(1 - y\right)\ln y,$$

where $y \in (0, 1)$. and we want to find

$$y^* = \text{argmax } g(y) = \text{argmax}\left[-\frac{1}{a}\ln\left(1 - y\right)\ln y\right].$$

First, we show that $g(y)$ is convex by examining the sign of its second derivative. Upon differentiating, we find

$$\frac{d}{dy}g(y) = -\frac{1}{a}\left[\frac{\ln(1-y)}{y} - \frac{\ln y}{1-y}\right]. \tag{6}$$

The second derivative is

$$\frac{d^2}{dy^2}g(y) = \frac{1}{a}\left[\frac{\ln(1-y)}{y^2} + \frac{2}{y(1-y)} + \frac{\ln y}{(1-y)^2}\right].$$

Consider the function $h(y) = 1 - y + y\ln y$. We find $h'(y) = \ln y < 0$ for $y \in (0,1)$. Therefore $h(y)$ is strictly decreasing over $(0,1)$. Since $h(1) = 0$, $h(y) > 0 \; \forall \; y \in (0,1)$. Also, since $y(1-y)^2 > 0 \; \forall \; y \in (0,1)$, we can divide by that quantity to obtain

$$\frac{\ln y}{(1-y)^2} + \frac{1}{y(1-y)} > 0 \; \forall \; y \in (0,1).$$

Similarly, it can be shown that

$$\frac{\ln(1-y)}{y^2} + \frac{1}{y(1-y)} > 0 \; \forall \; y \in (0,1).$$

Therefore $g(y)$ is convex over $y \in (0,1)$.

Given convexity, argmax $g(y)$ can be obtained by solving $g'(y) = 0$ for y. From

$$\frac{d}{dy}g(y) = -\frac{1}{a}\left[\frac{\ln(1-y)}{y} - \frac{\ln y}{1-y}\right] = 0, \tag{7}$$

it follows that

$$(1-y^*)\ln(1-y^*) = y^*\ln y^*$$

from which it is clear that $y^* = 1 - y^*$, which in turn results in $y^* = 0.5$. Thus, since $a = 2T_{\mathrm{A}}(N_A - 1)$ and $x = 1/\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]$, we find the optimal value of $\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]$ is given by

$$\mathrm{E}^*\left[\tilde{t}_{\mathrm{IA}}\right] = -\frac{2T_{\mathrm{A}}(N_A-1)}{\ln 0.5} = \frac{2T_{\mathrm{A}}(N_A-1)}{\ln 2}. \tag{8}$$

Upon solving (3) for T_S, we find the minimal value of T_S required to scan all advertisers while satisfying the target probability of success is

$$T_S = \mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]\frac{\ln(\text{target } P\{\mathrm{SJF}\})}{\ln 0.5} \tag{9}$$

To the best of our knowledge the results shown in (8) and (9) have not been previously presented in the literature.

III. Preliminary Simulation Results and Analysis

In this section we present simulation results based on the parameter settings derived in section II. We developed a detailed simulation of the scanning system in Simpy, a discrete event simulation package, using Python 2.7. We simulated systems having batch sizes of 100, 200, 400, 800 and 1600 with a target $P\{\mathrm{SJF}\} \leq 10^{-5}$. For each set of parameter settings we ran 30 different simulation runs, each of which simulated the scanning process for 1 million advertisers. At a failure rate less than 10^{-5}, there should be less than 10 out of 1 million SJFs. By doing about 30 samples of 10^6 SJs, it should be possible to determine the actual $P\{\mathrm{SJF}\}$ to a reasonable degree of accuracy.

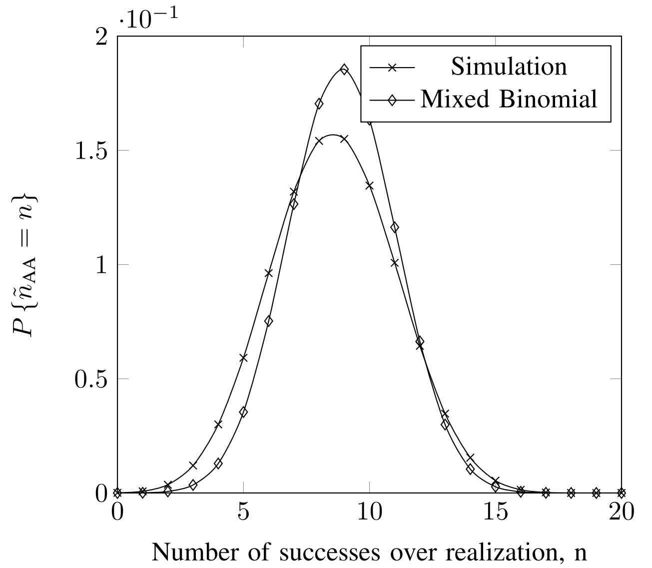

Fig. 3. Comparison of probability mass functions of number of successes for a batch size of 800 advertisers with a scanning period of 14.9 s and an average advertisement cycle of 0.8668 μs based on simulation and mixed binomial distributions.

We now turn to the selection of the values of $\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]$ and T_S. First we computed $\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]$ and T_S using (8) and (9). For example, for the case of $N_A = 800$ and with $T_{\mathrm{A}} = 376$ μs, $\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right] = 0.8668$ and $T_S = 14.398$. There would then be an average of $T_S/\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right] = 16.61$ advertising cycles within a scanning period. Analysis shows that in this case 61 % of the advertisers would generate 17 advertisements, and 39 % would generate only 16 advertisements, thus giving those a lower probability of success. Given that the random period is uniform over (0,10) ms, a maximum-length advertising cycle would be 0.8718 s. In order to ensure that all advertisers get at least 17 advertisements, the scanning period is increased to 17 maximum-length advertising cycles so that $T_S = 14.82$ s, which we rounded up to 14.9. At this setting, a little over 81.1 % of the advertisers would have 17 advertising cycles and the other 18.9 % would have 18.

Fig. 3 shows the probability mass function (PMF) for the number of successful scans over the scanning period for $N_A = 800$, $\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right] = 0.8668$, $T_S = 14.90$, and $\tilde{n}_{\mathrm{AA}}$ is defined as the number of successful scans of an arbitrary advertiser over a scanning period. Analysis of the simulation results shows that the average number of successes per advertiser, that is $\mathrm{E}\left[\tilde{n}_{\mathrm{AA}}\right]$, is 8.5923 and $P\{\mathcal{S}_{\mathrm{AA}}\} = 0.4999$, which is exactly according to the the design specification.

Also shown in Fig. 3 is the PMF for a mixed binomial distribution with 81.1 % of the advertisers having 17 advertising cycles, the other 18.9 % having 18 and probability of success set at 0.5 for which the average number of successes is 8.5945. Thus, the results from the simulation show $P\{\mathcal{S}_{\mathrm{AA}}\}$ and the average number of successfully read advertisements for an arbitrary advertiser are exactly as expected. On the other hand, it is readily seen by observation of Fig. 3 that the probability masses obtained from the simulation are significantly larger than those of the mixed binomial RV in the region $\tilde{n}_{\mathrm{AA}} < 7$ with corresponding lower probability masses in the region $7 < \tilde{n}_{\mathrm{AA}} < 12$.

Thus, it is seen that the simulations show that the actual scanning process is likely to have a higher probability of having fewer successes than would be indicated under the assumption of Poisson arrivals, which would result in the mixed binomial distribution shown. In fact, as illustrated in Table I, the simulation indicates $P\{\mathrm{SJF}\} = 7.45 \times 10^{-5}$ with a standard deviation of 8.55×10^{-7}. Thus, the observed average $P\{\mathrm{SJF}\}$ is about 7 times the target of 10^{-5} with the standard deviation indicating a very good estimate. In addition, the analytical formula yields $P\{\mathrm{SJF}\} = 4.54 \times 10^{-6}$ so that the average $P\{\mathrm{SJF}\}$ observed from simulations is about 16 times larger than would be expected if collisions were independent.

TABLE I
$P\{\mathrm{SJF}\}$ BASED ON PARAMETER SETTINGS ASSUMING INDEPENDENT COLLISIONS

N_{A}	$\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]$	T_{S} (s)	$P\{\mathrm{SJF}\}$	CI [95 %]
100	0.107	1.91	4.04e-05	± 1.377e-06
200	0.216	3.80	5.55e-05	± 9.119e-07
400	0.433	7.50	6.52e-05	± 9.485e-07
800	0.867	14.90	7.45e-05	± 6.555e-07
1600	1.740	29.60	7.74e-05	± 4.542e-07

Results for batch sizes of 100 and 1600 are shown in Figures 4 and 5, and it is readily seen from those figures that the accuracy of the analytical results is approximately the same for a wide variety of batch sizes. It is also seen that the bias towards higher probabilities of small numbers of successes is more pronounced as batch size increases. It is therefore of interest to examine the nature of the dependence among the collisions to which we now turn.

IV. Dependence Among Collisions

In this section we address the issue of dependence among collisions. We first address the issue of correlation of collision events through direct probabilistic analysis in subsection IV-A. In that subsection we show, by deriving probabilities, that the probability of additional collisions of a given advertiser with the tagged advertiser given an initial collision with the tagged advertiser is significantly higher than that of an arbitrary advertisement.

In subsection IV-B, we pursue the issue of correlation of collision events via simulation. More specifically, we define events involving strings of collisions of various lengths and measure the probability of extending the string by one collision for each of the string lengths. The results of that subsection indicate that the probabilities increase as string length increases, but then seem to converge as string length increases.

A. Analytical Approach to Collision Dependence

To begin the discussion, it is noted that while initial transmissions of advertisements over a scanning interval occur at random points in time over an interadvertisement period, the times of future transmissions are heavily correlated to the initial transmission times. In this subsection, we examine the effect that correlation has on collision probabilities. We denote the tagged advertiser as advertiser A. Suppose advertiser A begins an advertisement at time t_0 and an advertisement from advertiser B overlaps with that of A. It then follows that the advertisement of B started at a random time, say $\tilde{t}_{B,0}$, sometime during the interval $(t_0 - T_{\mathrm{A}}, t_0 + T_{\mathrm{A}})$. The following transmissions of A and B then occur at times $\tilde{t}_{A,1} = t_0 + T_{\mathrm{AI}} + \tilde{t}_{AI,A}$ and $\tilde{t}_{B,1} = \tilde{t}_{B,0} + T_{\mathrm{AI}} + \tilde{t}_{AI,B}$, where $\tilde{t}_{AI,A}$

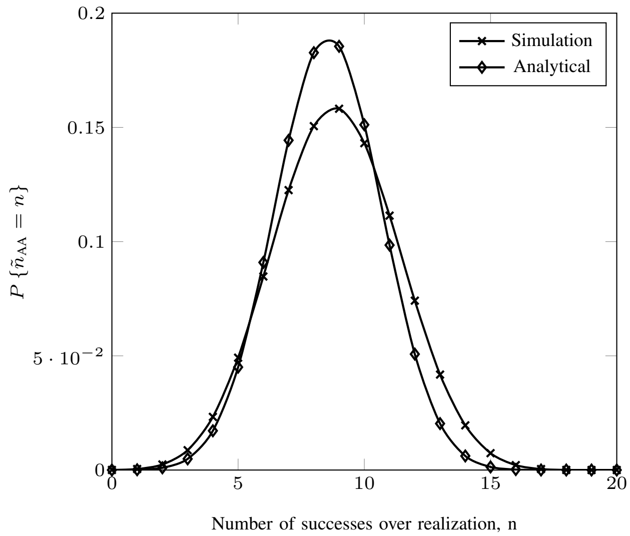

Fig. 4. Comparison of probability mass functions of number of successes for a batch size of 100 advertisers with a scanning period of 1.91 s and an average advertisement cycle of 0.107 s based on simulation and mixed binomial distributions.

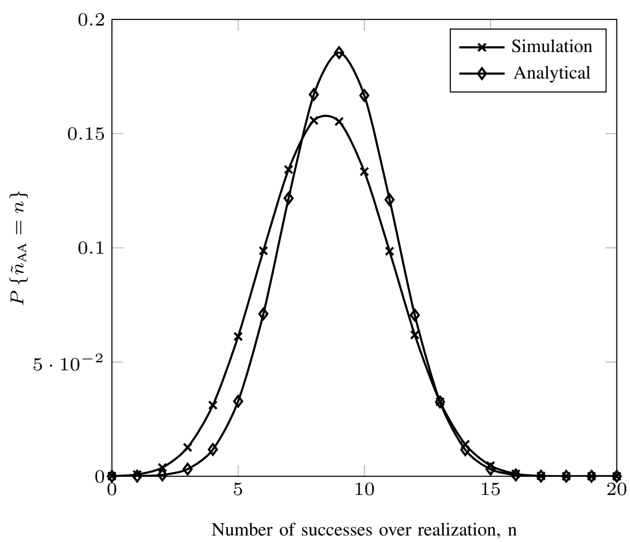

Fig. 5. Comparison of probability mass functions of number of successes for a batch size of 1600 advertisers with a scanning period of 29.6 s and an average advertisement cycle of 1.74 s based on simulation and mixed binomial distributions.

and $\tilde{t}_{AI,B}$ denote the random parts of the interadvertisement times. The difference between the next advertisements is then

$$\begin{aligned}\tilde{t}_{A,1}-\tilde{t}_{B,1} &= t_0+T_{\mathrm{AI}}+\tilde{t}_{AI,A}-\left[\tilde{t}_{B,0}+T_{\mathrm{AI}}+\tilde{t}_{AI,B}\right] \\ &= t_0-\tilde{t}_{B,0}+\tilde{t}_{AI,A}-\tilde{t}_{AI,B}, \end{aligned} \quad (10)$$

which we note is independent of T_{AI}. A second collision will occur if $-T_A \leq \tilde{t}_{A,1}-\tilde{t}_{B,1} \leq T_A$.

To facilitate further discussion, we define the following events:

- $\mathcal{C}_n$ Event of nth repeated collision of the tagged advertiser. For example, $\mathcal{C}_0$ is the tagged advertiser's initial collision and $\mathcal{C}_1$ is the event of tagged advertiser's second consecutive collision, which would occur on its first advertisement after its first collision.
- $\mathcal{N}_n$ Event of a collision that occurs due to an advertiser whose first collision with the tagged advertiser occurs at the nth advertisement epoch of the tagged advertiser.
- $\mathcal{R}_n$ Event of a repeated collision that occurs in the n advertisement epoch of the tagged advertiser. In particular, $\mathcal{R}_{\mathcal{B},n}$, will denote a repeated collision of advertiser B at epoch n, $n=1,2$, and $\mathcal{R}_{\mathcal{C},2}$ will denote a repeated collision of advertiser C at epoch 2.

Since the initial advertisements occur at random independent times, the random variable $t_0-\tilde{t}_{B,0}$ can be assumed to be uniformly distributed over $(-T_{\mathrm{A}},T_{\mathrm{A}})$, and $\tilde{t}_{AI,A}$ and $\tilde{t}_{AI,B}$ are each uniformly distributed over $(0,T_{\mathrm{D}})$ so that $-\tilde{t}_{AI,B}$ is uniformly distributed over $(-T_{\mathrm{D}},0)$. In addition the three random variables just mentioned are independent so that the distribution of $\tilde{t}_{A,1}-\tilde{t}_{B,1}$ is the threefold convolution of uniformly distributed random variables.

By using Laplace-Stieltjes transform techniques, it is readily found that the part of the probability density function (PDF) of $\tilde{t}_{A,1}-\tilde{t}_{B,1}$ where the event $\mathcal{R}_{\mathcal{B},1}$ can occur, $(-T_{\mathrm{A}},T_{\mathrm{A}})$, is given by

$$f_{\tilde{t}_{A,1}-\tilde{t}_{B,1}}(t)=\frac{1}{2T_D^2}\left(2T_D-T_A-\frac{1}{T_A}t^2\right) \quad (11)$$

Upon integrating over the interval $(-T_{\mathrm{A}},T_{\mathrm{A}})$, we then find

$$P\left\{\mathcal{R}_{\mathcal{B},1}|\mathcal{C}_0\right\}=2\frac{T_{\mathrm{A}}}{T_{\mathrm{D}}}\left(1-\frac{2}{3}\frac{T_{\mathrm{A}}}{T_{\mathrm{D}}}\right)=0.0733, \quad (12)$$

where the given numerical probability is based on the standard values of $T_{\mathrm{A}}=0.376$ ms and $T_{\mathrm{D}}=10$ ms as specified earlier. Note that $P\left\{\mathcal{R}_{\mathcal{B},1}|\mathcal{C}_0\right\}$ does not depend upon any other parameters; with the standard parameter settings, it is always true that $P\left\{\mathcal{R}_{\mathcal{B},1}|\mathcal{C}_0\right\}=0.0733$.

The distribution of the difference in arrival times given a second collision is obtained by normalizing (11) using (12). The result, both in term of symbols and numbers representing standard parameter values, is as follows:

$$\begin{aligned}f_{\tilde{t}_{A,1}-\tilde{t}_{B,1}|\mathcal{R}_{\mathcal{B},1}}(t) &= \frac{3T_A\left(2T_D-T_{\mathrm{A}}\right)-3t^2}{4T_{\mathrm{A}}^2\left(3T_D-2T_{\mathrm{A}}\right)} \\ &= 1.338-0.1814t^2, t\in(-0.376,0.376)\end{aligned} \quad (13)$$

A careful examination of the above conditional distribution reveals that it is very close to a uniform distribution over $(-T_{\mathrm{A}},T_{\mathrm{A}})$, but the mass is slightly more concentrated to the center. Thus, we would expect that the probability of a third collision given a second would be very nearly the same as, but slightly larger, than $P\left\{\mathcal{R}_{\mathcal{B},1}|\mathcal{C}_0\right\}$. A detailed analysis reveals that $P\left\{\mathcal{R}_{\mathcal{B},2}|\mathcal{R}_{\mathcal{B},1}\right\}=0.0736$ versus $P\left\{\mathcal{R}_{\mathcal{B},1}|\mathcal{C}_0\right\}=0.0733$.

To put the above numbers in perspective, the probability of a collision due to a single independent advertiser is $2T_A/\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]$, which turns out to be 8.675×10^{-4} for a group of $N_A=800$ with $\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]=0.8668$. The conditional probability of a repeat collision with the same advertiser is about 85 times as large as the probability of collision with a single random advertiser.

Since the event $\mathcal{C}_1|\mathcal{C}_0\mathcal{N}_1$ is certain, upon conditioning on the occurrence of $\mathcal{N}_1$,

$$P\left\{\mathcal{C}_1|\mathcal{C}_0\right\}=P\left\{\mathcal{N}_1|\mathcal{C}_0\right\}+P\left\{\mathcal{C}_1|\mathcal{C}_0\bar{\mathcal{N}_1}\right\}P\left\{\bar{\mathcal{N}_1}|\mathcal{C}_0\right\} \quad (14)$$

Under the assumption of independence between repeat collision events and those resulting from the general population, we find $P\left\{\mathcal{C}_1|\mathcal{C}_0\bar{\mathcal{N}_1}\right\}=P\left\{\mathcal{R}_{B,1}|\mathcal{C}_0\right\}$. In addition, for $\mathcal{N}_1$, there are N_A-2 advertisers in the general population. Thus with the optimal settings,

$$P\left\{\mathcal{N}_1|\mathcal{C}_0\right\}=1-e^{-\frac{2T_{\mathrm{A}}(N_{\mathrm{A}}-1-1)}{\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]}}=1-0.5e^{\frac{2T_{\mathrm{A}}}{\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]}} \quad (15)$$

Numerically, for the case of 800 advertisers, $P\left\{\mathcal{N}_1|\mathcal{C}_0\right\}=0.4996$. Since we have shown that $\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]$ is very close to linear in the number of advertisers, the result will be very close for any value of N_A. Upon substituting numerical values into (14), we find

$$P\left\{\mathcal{C}_1|\mathcal{C}_0\right\}=0.4996+0.0733(1-0.4996)=0.5363 \quad (16)$$

Given $\mathcal{C}_1$, the possible events are $\mathcal{R}_{B,1}\bar{\mathcal{N}_1}$, $\bar{\mathcal{R}}_{B,1}\mathcal{N}_1$, and $\mathcal{R}_{B,1}\mathcal{N}_1$, for which the probabilities are obtained by various multiplications of the previously computed event probabilities as a result of independence. The resulting probabilities are $0.0367, 0.4630$, and 0.0366 for $\mathcal{R}_{B,1}\bar{\mathcal{N}_1}$, $\bar{\mathcal{R}}_{B,1}\mathcal{N}_1$, and $\mathcal{R}_{B,1}\mathcal{N}_1$, respectively, and their conditional probabilities given $\mathcal{C}_1$ are 0.0684, 0.8633, and 0.0683.

We now begin the development of $P\left\{\mathcal{C}_2|\mathcal{C}_1\right\}$, which we will accomplish by conditioning on the 3 outcomes of $\mathcal{C}_1$. First we note that the event $\mathcal{C}_2|\mathcal{C}_1\mathcal{N}_2$ is certain. In addition, $\mathcal{N}_2$ is dependent upon $\mathcal{N}_1$ but not $\mathcal{R}_{B,1}$ so that

$$P\left\{\mathcal{C}_2|\mathcal{R}_{B,1}\bar{\mathcal{N}_1}\right\}=P\left\{\mathcal{N}_2|\bar{\mathcal{N}_1}\right\}+P\left\{\mathcal{R}_{B,2}|\mathcal{R}_{B,1}\right\}P\left\{\bar{\mathcal{N}_2}|\bar{\mathcal{N}_1}\right\} \quad (17)$$

But, $P\left\{\mathcal{N}_2|\bar{\mathcal{N}_1}\right\}=P\left\{\mathcal{N}_1|\mathcal{C}_0\right\}$, which is given by (15), and $P\left\{\mathcal{R}_{B,2}|\mathcal{R}_{B,1}\right\}$ is specified above as 0.0736. Thus, we have numerically, $P\left\{\mathcal{C}_2|\mathcal{R}_{B,1}\bar{\mathcal{N}_1}\right\}=0.53643$.

Next, we consider $\bar{\mathcal{R}}_{B,1}\mathcal{N}_1$. In this case, $\mathcal{C}_2$ occurs if the event $\mathcal{N}_2\cup\mathcal{R}_{B,2}\cup\mathcal{R}_{C,2}$ occurs. Again, the event $\mathcal{C}_2|\mathcal{C}_1\mathcal{N}_2$ is certain. $\mathcal{N}_2$ is dependent upon $\mathcal{N}_1$ but not $\mathcal{R}_{B,1}$. The events $\mathcal{R}_{B,2}$ and $\mathcal{R}_{C,2}$ are not independent because they are both dependent upon $\tilde{t}_{A,2}$. In fact, $P\left\{\mathcal{R}_{B,2}\mathcal{R}_{C,2}\right\}>P\left\{\mathcal{R}_{B,2}\right\}P\left\{\mathcal{R}_{C,2}\right\}$ due to the positive correlation of the events. On the other hand, an exact analysis shows the impact of assuming independence of these particular events has minute effects on the overall result. Thus, under the independence assumption, we have

$$P\left\{\mathcal{C}_2|\bar{\mathcal{R}}_{B,1}\mathcal{N}_1\right\} = P\left\{\mathcal{N}_2|\mathcal{N}_1\right\} + P\left\{\mathcal{R}_{B,2}\cup\mathcal{R}_{C,2}|\bar{\mathcal{R}}_{B,1}\mathcal{N}_1\right\} P\left\{\bar{\mathcal{N}_2}|\mathcal{N}_1\right\} \quad (18)$$

and

$$P\left\{\mathcal{R}_{B,2}\cup\mathcal{R}_{C,2}|\bar{\mathcal{R}}_{B,1}\mathcal{N}_1\right\} = P\left\{\mathcal{R}_{B,2}|\bar{\mathcal{R}}_{B,1}\right\} + P\left\{\mathcal{R}_{C,2}|\mathcal{N}_1\right\} - P\left\{\mathcal{R}_{B,2}|\bar{\mathcal{R}}_{B,1}\right\} P\left\{\mathcal{R}_{C,2}|\mathcal{N}_1\right\} \quad (19)$$

At this point, we have $P\left\{\mathcal{R}_{C,2}|\mathcal{N}_1\right\} = P\left\{\mathcal{R}_{B,1}|\mathcal{C}_0\right\}$ and all probabilities except $P\left\{\mathcal{R}_{\mathcal{B},2}|\bar{\mathcal{R}}_{\mathcal{B},1}\right\}$ and $P\left\{\mathcal{N}_2|\mathcal{N}_1\right\}$. By proceeding along the same lines as before, we find the difference in the advertisement times of the second advertisement after the initial collision is the sum of two independent $U\left(-T_D,0\right)$, two independent $U\left(0,T_D\right)$ and one $U\left(-T_A,T_A\right)$ random variables. The part of the PDF over the interval in $\mathcal{R}_{B,2}$ is found to be

$$\begin{aligned} f_{\tilde{t}_{A,2}-\tilde{t}_{B,2}}(t) &= \frac{1}{T_A T_D^4}\left[a_4 t^4 + a_2 t^2 + a_0\right], \\ a_4 &= \frac{1}{8}, \quad a_2 = \frac{3T_A^2 - 4T_A T_D}{4}, \\ a_0 &= \frac{2T_D^3 T_A - T_D T_A^3}{3} + \frac{T_A^4}{8} \end{aligned} \quad (20)$$

Upon integrating the above PDF over $(-T_{\mathrm{A}},T_{\mathrm{A}})$, we find

$$P\left\{\mathcal{R}_{\mathcal{B},2}|\mathcal{C}_0\right\} = \frac{4T_A}{T_D}\left[\frac{1}{3} - \frac{1}{3}\left(\frac{T_A}{T_D}\right)^2 + \frac{1}{5}\left(\frac{T_A}{T_D}\right)^3\right]$$

Numerically, $P\left\{\mathcal{R}_{\mathcal{B},2}|\mathcal{C}_0\right\} = 0.0501$. Thus, it is seen again that conditional probabilities of collision given past collisions are significantly higher than would be expected from independent trials.

Now, since we know $P\left\{\mathcal{R}_{\mathcal{B},2}|\mathcal{R}_{\mathcal{B},1}\right\}$ and $P\left\{\mathcal{R}_{\mathcal{B},1}|\mathcal{C}_0\right\}$, we can find $P\left\{\mathcal{R}_{\mathcal{B},2}|\bar{\mathcal{R}}_{\mathcal{B},1}\right\}$ from

$$P\left\{\mathcal{R}_{\mathcal{B},2}|\mathcal{C}_0\right\} = P\left\{\mathcal{R}_{\mathcal{B},2}|\mathcal{R}_{\mathcal{B},1}\right\} P\left\{\mathcal{R}_{\mathcal{B},1}|\mathcal{C}_0\right\} + P\left\{\mathcal{R}_{\mathcal{B},2}|\bar{\mathcal{R}}_{\mathcal{B},1}\right\} P\left\{\bar{\mathcal{R}}_{\mathcal{B},1}|\mathcal{C}_0\right\} \quad (21)$$

Numerically, $P\left\{\mathcal{R}_{\mathcal{B},2}|\bar{\mathcal{R}}_{\mathcal{B},1}\right\} = 0.0482$. Proceeding as in (15) gives

$$P\left\{\mathcal{N}_2|\mathcal{N}_1\right\} = 1 - e^{-\frac{2T_{\mathrm{A}}\left(N_{\mathrm{A}}-1-2\right)}{\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]}} = 1 - 0.5e^{\frac{4T_{\mathrm{A}}}{\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]}} \quad (22)$$

Numerically, $P\left\{\mathcal{N}_2|\mathcal{N}_1\right\} = 0.4991$. Thus, from (18), (19), and (21) we have numerically

$$P\left\{\mathcal{C}_2|\bar{\mathcal{R}}_{B,1}\mathcal{N}_1\right\} = 0.5582. \quad (23)$$

Finally, we consider the case of $\mathcal{R}_{B,1}\mathcal{N}_1$. In this case $\mathcal{C}_2$ also occurs if the event $\mathcal{N}_2\cup\mathcal{R}_{B,2}\cup\mathcal{R}_{C,2}$ occurs. A detailed analysis of this case has been carried out including that of dependent events. The analysis parallels that of $\bar{\mathcal{R}}_{B,1}\mathcal{N}_1$, the primary difference being that the event $\mathcal{R}_{B,2}|\mathcal{R}_{B,1}$ has a higher probability than the event $\mathcal{R}_{B,2}|\bar{\mathcal{R}}_{B,1}$. The analysis showed that the dependence again has very minor impact upon the results. The result based on the assumption of independence is then $P\left\{\mathcal{C}_2|\mathcal{R}_{B,1}\mathcal{N}_1\right\} = 0.5699$. Upon combining the conditional probabilities, the result is

$$P\left\{\mathcal{C}_2|\mathcal{C}_1\right\} = 0.5575, \quad (24)$$

which will be slightly higher than the actual value due to the independence assumption.

Thus, it is seen that $P\left\{\mathcal{C}_2|\mathcal{C}_1\right\} > P\left\{\mathcal{C}_1|\mathcal{C}_0\right\}$. In addition, comparing $P\left\{\mathcal{R}_{B,1}|\mathcal{C}_0\right\} = 0.0733$ to $P\left\{\mathcal{R}_{B,2}|\mathcal{C}_0\right\} = 0.0501$, we see that the unconditional probability of repeated collisions decreases significantly as the number advertisements beyond the first collision increases. Thus, it would be expected that the impact of a specific advertiser on the overall probability of continued collisions would decrease over time with the result that the $P\left\{\mathcal{C}_n|\mathcal{C}_{n-1}\right\}$ would tend to converge with increasing n. The nature of the convergence process is still under study, but the observation is used in the next section to design systems. Meanwhile it is noted that all of the formulas involving the probabilities of repeated collisions were checked using specialized Python simulations. We now turn to the examination of repeated collisions via simulation analysis.

B. Simulation Approach to Collision Dependence

In this section, we discuss our approach to investigating dependence among successive collisions via simulation. In our detailed simulation program, we collected traces of the results of each advertisement for each advertiser in the form $[r_1, r_2, r_3, \ldots]$, $r_n \in \{s, f\}$, where f (s) represents a collision (no collision). We then did a simple analysis of the resulting lists to determine the proportion of collision events that follow collision events; for example, if $r_n = f$ and $r_{n+1} = f$, this counts as one collision event that follows a collision event. The count obtained was normalized to the number of transitions from collision events. We used the result as an approximation of $P\left\{\mathcal{R}_{B,1}|\mathcal{C}_0\right\}$. Note that what is being reported is the probability that an arbitrary collision event is followed by a collision event, whereas $P\left\{\mathcal{R}_{B,1}|\mathcal{C}_0\right\}$ is the probability that the initial collision event is followed by a collision event.

Similarly, we compute the probability that two collision events are followed by a collision event. For example, if $[r_n, r_{n+1}] = [f, f]$ and $[r_{n+1}, r_{n+2}] = [f, f]$, this counts as one event. This is used as an approximation of $P\left\{\mathcal{R}_{B,2}|\mathcal{R}_{B,1}\right\}$. In the case we look for repeated collisions based only on r_n or only on $[r_n, r_{n+1}]$, we say the memory length is 1 or 2, respectively. The results of 600 replications of 180,000 s simulations of only two advertisers gave the mean and 95 % confidence intervals of the two conditional events as 0.0730 ± 0.007 and 0.0714 ± 0.0214, respectively, where the large confidence interval of the latter is the result of scarcity of events. Note that the 95 % confidence intervals (CI), which are based on the mean $\pm$ 1.96 standard deviations, cover the theoretical values of $P\left\{\mathcal{R}_{B,1}|\mathcal{C}_0\right\}$ and $P\left\{\mathcal{R}_{B,2}|\mathcal{R}_{B,1}\right\}$, but the confidence intervals are quite large even though the runs are fairly long because the events are fairly rare when only two advertisers are present. Nonetheless, the results do support the result that repeated collisions of the same two advertisers are much more likely than arbitrary random collisions. Notice that the standard deviations increase as the memory length increases; this is a result of the fact that for any given simulation period, the number of events upon which the mean value is based decreases with memory length.

TABLE II
CONDITIONAL PROBABILITIES OF COLLISION FOR A RANGE OF GROUP SIZES AND MEMORY LENGTHS

Memory Length	Mean Collision Return Probability Number of Advertisers				
	100	200	400	800	1600
1	0.525	0.525	0.525	0.526	0.527
2	0.538	0.540	0.541	0.542	0.543
3	0.547	0.552	0.554	0.555	0.556
4	0.554	0.563	0.565	0.566	0.567
5	0.559	0.572	0.574	0.575	0.577
6	0.565	0.580	0.582	0.584	0.586
7	0.571	0.587	0.590	0.592	0.593
8	0.571	0.593	0.597	0.599	0.600

Continuing on, we examine the rates of repeated collisions via detailed simulations for $N_A \in \{100, 200, 400, 800, 1600\}$ with $\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]$ set optimally and memory length as a parameter. The results, shown in Table II, are used as a proxy for $P\left\{\mathcal{C}_n|\mathcal{C}_{n-1}, \mathcal{C}_{n-2}, \ldots, \mathcal{C}_0\right\}$ as defined in the previous section. Discussion of specific confidence intervals is omitted here as they are all sufficiently small that they would affect the accuracy of the estimates only in the third decimal point and do not have any direct effect on the substance of this particular discussion.

As can be seen from the table, $P\left\{\mathcal{C}_n|\mathcal{C}_{n-1}, \mathcal{C}_{n-2}, \ldots, \mathcal{C}_0\right\}$ is, in fact, increasing as the memory length grows as conjectured in the previous subsection. It is also again seen that the standard deviations increase with memory length for reasons explained above.

V. Accounting for Correlation of Collisions

Given the observations of the previous section, it is clear that correlation in the collision process has a major impact upon $P\{\mathrm{SJF}\}$. Bearing in mind that a SJF occurs only if all of an advertiser's advertisements collide, it is clear that SJFs are the result of a continuous string of collisions, including the first advertisement of the scanning period. In addition, at the time scanning begins, the number of collisions that have already occurred in the current string is not known. Given that

$$P\left\{\mathcal{C}_n, \mathcal{C}_{n-1}, \ldots, \mathcal{C}_0\right\} = P\left\{\mathcal{C}_n|\mathcal{C}_{n-1}\mathcal{C}_{n-2}, \ldots, \mathcal{C}_0\right\} P\left\{\mathcal{C}_{n-1}, \mathcal{C}_{n-2}, \ldots, \mathcal{C}_0\right\} \quad (25)$$

it is clear that the scan time to achieve a given $P\{\mathrm{SJF}\}$ target must be determined based on the probability of repeated collisions because SJFs are the result of repeated collisions.

Based on the numerical results of the previous section, it appears that $P\left\{\mathcal{C}_n|\ \mathcal{C}_{n-1}\right\}$ converges to approximately 0.6 as n increases, thus we assume the value of 0.6 to illustrate the idea. We revisit the analysis of $P\{\mathrm{SJF}\}$ as illustrated in (3) and suppose that there is a negative bias of B in the probability of success that future advertisements from an advertiser whose first advertisement of the observation period collides. Then from (1), we find

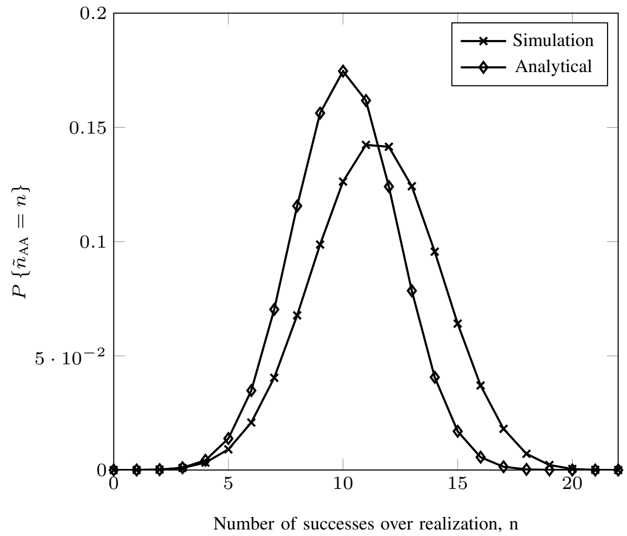

Fig. 6. Comparison of probability mass functions of number of successes for a batch size of 100 advertisers with a scanning period of 2.59 s and an average advertisement cycle of 0.1245 s based on simulation and mixed binomial distributions.

$$P\left\{\mathcal{S}_{\mathrm{AA}}^{\mathrm{bias}}\right\} \approx e^{-2\rho_{\mathrm{A}}(N_A-1)} - B. \quad (26)$$

Thus

$$P\left\{\mathrm{SJF}^{\mathrm{bias}}\right\} = \left[1 - \left(e^{-2\frac{T_{\mathrm{A}}(N_A-1)}{\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]}} - B\right)\right]^{\frac{T_S}{\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]}}. \quad (27)$$

Then upon following the same analysis approach as was followed in section II, we find $y* = \frac{1+B}{2}$ so that

$$\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}^{\mathrm{bias}}\right] = -\frac{2T_A\left(N_A-1\right)}{\ln\left(\frac{1+B}{2}\right)} = -\frac{2T_A\left(N_A-1\right)}{\ln\left(.55\right)} \quad (28)$$

and

$$T_S^{\mathrm{bias}} \geq \mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right] \frac{\ln\left(\text{target } P\{\mathrm{SJF}\}\right)}{\ln 0.55} \quad (29)$$

Table III shows the resulting $P\{\mathrm{SJF}\}$s for $N_A \in \{100, 200, 400, 800, 1600\}$ with $\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]$s set optimally and required scan periods based on independent collisions with the target $P\{\mathrm{SJF}\}$ set at 10^{-5}. Confidence intervals for the dependent case are looser than those obtained for the

TABLE III
$P\{\mathrm{SJF}\}$ BASED ON PARAMETER SETTINGS ASSUMING DEPENDENT COLLISIONS

N_{A}	$\mathrm{E}\left[\tilde{t}_{\mathrm{IA}}\right]$	T_{S} (s)	$P\{\mathrm{SJF}\}$	CI [95 %]
100	0.125	2.59	1.83e-06	± 2.064e-06
200	0.250	5.11	3.77e-06	± 4.358e-06
400	0.502	10.14	3.50e-06	± 4.045e-06
800	1.005	20.20	4.73e-06	± 3.951e-06
1600	2.011	40.32	4.87e-06	± 3.481e-06

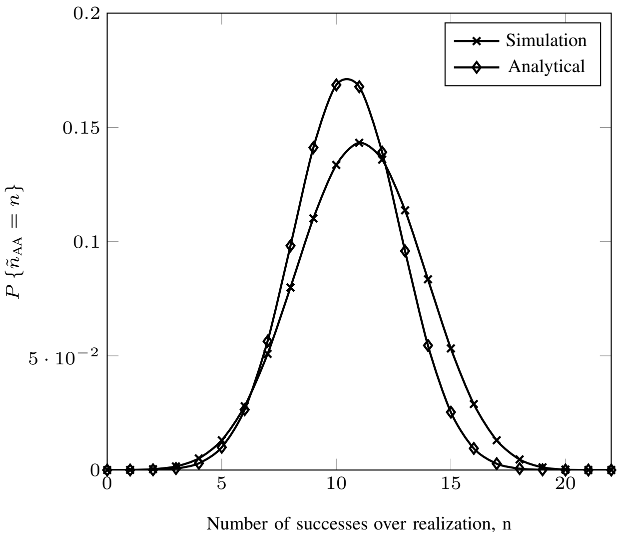

Fig. 7. Comparison of probability mass functions of number of successes for a batch size of 1600 advertisers with a scanning period of 40.32 s and an average advertisement cycle of 2.011 s based on simulation and mixed binomial distributions.

independent case due to the lower number of failure events. The upper end of the 95 % confidence interval, labeled UCI, which is given by the mean plus 1.96 standard deviations, for the case of dependent collisions is also given in the table.

From Table I it is seen that when the scanning period is designed under the assumption of independent collisions, the target $P\{\mathrm{SJF}\}$ is missed by at least a factor of 4 even without considering confidence intervals. On the other hand, from Table III shows that if dependence among collisions is taken into account the resulting $P\{\mathrm{SJF}\}$ always meets the target at the 95 % confidence level.

Figures 6 and 7 show comparisons probability mass functions of number of successes for batch sizes 100 and 1600 advertisers with interadvertisement times and scanning periods based on dependent collisions. From these graphs it can be seen that the simulation curves track much more closely with the analytical curves at the low end of the distribution so that the accuracy of the prediction of the probability of scanning job failure is greatly improved. Indeed rather than under predicting the probability of failure the modified approach tends to overpredict the probability of failure thereby resulting in a conservative estimate, which is the desired result.

VI. Conclusions

We have presented here a detailed analysis and design of a BLE scanning system wherein the objective is to choose optimal interadvertisement times and minimum scanning times required to meet a prescribed target probability of failing to successfully scan all advertisers within the scanning period. To the best of our knowledge, this is the first work to do a detailed analysis of the impact of dependence in the advertising process upon the collision probability and to show how this dependence affects the probability of successfully scanning all members of a group.

References

[1] B. SIG, *BLUETOOTH SPECIFICATION Version 5.0 — Vol 6, Part A*, Dec. 2016.

[2] G. Shan and B. Roh, "Advertisement interval to minimize discovery time of whole ble advertisers," *IEEE Access*, vol. 6, pp. 17817–17825, 2018.

[3] G. Shan, B. Lee, S. Shin, and B. Roh, "Design and implementation of simulator for analysis of ble broadcast signal collision," in *2017 International Conference on Information Networking (ICOIN)*, pp. 448–452, Jan 2017.

[4] Jia Liu, Canfeng Chen, and Yan Ma, "Modeling and performance analysis of device discovery in Bluetooth Low Energy networks," in *2012 IEEE Global Communications Conference (GLOBECOM)*, pp. 1538–1543, Dec 2012.

[5] C. Julien, C. Liu, A. L. Murphy, and G. P. Picco, "Blend: Practical continuous neighbor discovery for Bluetooth Low Energy," in *2017 16th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 105–116, 2017.

[6] J. Liu, C. Chen, Y. Ma, and Y. Xu, "Adaptive device discovery in Bluetooth Low Energy networks," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, June 2013.

[7] J. Kim and K. Han, "Backoff scheme for crowded Bluetooth Low Energy networks," *IET Communications*, vol. 11, no. 4, pp. 548–557, 2017.

[8] M. Ghamari, E. Villeneuve, C. Soltanpur, J. Khangosstar, B. Janko, R. S. Sherratt, and W. Harwin, "Detailed examination of a packet collision model for bluetooth low energy advertising mode," *IEEE Access*, vol. 6, pp. 46066–46073, 2018.

[9] Yaw-Jen Lin, Heng-Shuen Chen, and Mei-Ju Su, "A cloud based bluetooth low energy tracking system for dementia patients," in *2015 Eighth International Conference on Mobile Computing and Ubiquitous Networking (ICMU)*, pp. 88–89, Jan 2015.

[10] Z. Yang, J. Schafer, and A. Ganz, "Disaster response: Victims' localization using bluetooth low energy sensors," in *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, pp. 1–4, April 2017.

[11] M. O. Al Kalaa and H. H. Refai, "Selection probability of data channels in Bluetooth Low Energy," in *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp. 148–152, Aug 2015.